

# Property Testing of Joint Distributions using Conditional Samples

Rishiraj Bhattacharyya  
NISER, Bhubaneswar  
India

rishiraj.bhattacharyya@gmail.com

Sourav Chakraborty  
Chennai Mathematical Institute,  
India

sourav@cmi.ac.in

February 7, 2017

## Abstract

In this paper, we present the first non-trivial property tester for joint probability distributions in the recently introduced conditional sampling model. The conditional sampling framework provides an oracle for a distribution  $\mu$  that takes as input a subset  $S$  of the domain  $\Omega$  and returns a sample from the distribution  $\mu$  conditioned on  $S$ . For a joint distribution of dimension  $n$ , we give a  $\tilde{O}(n^3)$ -query uniformity tester, a  $\tilde{O}(n^3)$ -query identity tester with a known distribution, and a  $\tilde{O}(n^6)$ -query tester for testing independence of marginals. Our technique involves an elegant chain rule which can be proved using basic techniques of probability theory, yet powerful enough to avoid the curse of dimensionality.

We also prove a sample complexity lower bound of  $\Omega(\sqrt[4]{n})$  for testing uniformity of a joint distribution when the tester is only allowed to condition independently on the marginals. Our technique involves novel relations between Hellinger distance and total variational distance, and may be of independent interest.

# 1 Introduction

PROPERTY TESTING OF DISTRIBUTIONS. The boom of Big Data Analytics has rejuvenated the well studied area of hypothesis testing over unknown distributions. In Computer Science, the study of this type of problems was initiated by Batu et al. [4] under the framework of “Property Testing.” In this framework, the “tester” draws independent samples from the distribution, and decides whether the distribution satisfies certain property  $\mathcal{P}$  (null hypothesis) or is far from any distribution that satisfies  $\mathcal{P}$  (alternate hypothesis).

Several properties of probability distributions have been studied in this framework. Testing whether the distribution is uniform [3, 12, 14], testing identity between two unknown distributions (taking samples from both the distributions) [4, 13], testing independence of marginals of product distributions [3], estimating entropy [2] are few of the numerous problems that have been studied in the literature. See [5] for a survey on results related to distribution testing.

Unfortunately, from the modern data analytics point of view, the classical framework of sampling yields impractical sample complexity. For example, testing if a distribution over a set of  $n$  elements is uniform requires  $\Omega(\sqrt{n})$  samples from the distribution. The other problems mentioned above have sample complexity at least this high, and in some cases almost linear in  $n$  [15, 16, 17].

CONDITIONAL SAMPLING. To remedy this situation, Chakraborty et. al. [7] and Canonne et. al. [6] proposed a different model, called conditional sampling, has emerged as a powerful tool for testing properties of probability distributions. In this model, the testers are allowed to sample according to the distribution conditioned on any specific subset of the domain. If the distribution,  $\mu$ , is over the domain  $\Sigma$ , the tester can submit any subset  $S \subseteq \Sigma$  and receive a sample  $i \in S$  with probability  $\mu(i) / \sum_{j \in S} \mu(j)$ , where  $\mu(i)$  is the probability of  $i$  occurring when a sample is drawn from the distribution  $\mu$ .

Chakraborty et. al. [7] proved that in the conditional sampling model, testing uniformity, testing identity to a known distribution, and testing any label-invariant property of distributions are easier than with the ordinary sampling model. Canonne et. al. [6] improved the query complexity and gave an algorithm for testing uniformity using  $\tilde{O}(\epsilon^{-2})$  conditional samples (conditioning on arbitrary subsets of size 2). Recently Falahatgar et. al. [9] showed that testing identity to a known distribution can also be done using  $\tilde{O}(1/\epsilon^2)$  conditional samples. They also showed that there exists an algorithm to test identity between two unknown distributions on  $\Sigma$  using  $\tilde{O}(\log \log |\Sigma| / \epsilon^5)$  conditional samples.

The sample complexity in the conditional sampling model depends on the structure of the condition, *i.e.*, the structure of the subsets (of the domain) on which the distribution is conditioned for drawing samples. Naturally, if there is no restriction on the condition, the tester can sample conditioned on arbitrary subset, the sample complexity improves. For example, in [6], authors presented an algorithm, for testing whether a distribution over  $\{1, \dots, n\}$  is uniform, with sample complexity  $\tilde{\Theta}(\epsilon^{-2})$  when conditioning on arbitrary subset of size 2. However, when the condition set was structured and restricted to intervals, they proved a lower bound of sample complexity  $\Omega\left(\frac{\log n}{\log \log n}\right)$ . Hence it is important to consider the plausible restrictions on the conditions, arising from the structure of the domain.

While [6] studied some of the restrictions of the conditions there are many more restrictions on the conditions, that arise from the structure of the domain and/or arise from other applications, which are yet to be studied. One such important case is when the domain is a Cartesian product of set and one is allowed to condition on Cartesian product of subsets, but not on arbitrary subset of the domain.

## Testing Joint Distributions: Subcube Conditioning

In practice, data are often multidimensional. In Cryptography, the keys are often defined over  $\{0, 1\}^n$ . Solutions to SAT formulae are over  $\{0, 1\}^n$  as well. On the other hand the Lottery Tickets are defined over  $[m]^n$  for some  $m \in \mathbb{N}$  (each ticket contains  $n$  numbers, each from the set  $[m]$ ). In fact, data analysts often get data of million dimensions (features). With higher dimension, comes the “curse of dimensionality”. The

sample complexity of the testers are often exponential in dimension [1], prohibiting practical applications. One can indeed be hopeful that using conditional sampling, the testers with practical complexity can be achieved. However, there is a caveat. As mentioned, the sample complexity heavily depends on the structure of the condition. In case of joint distributions, sampling conditioned on arbitrary subsets may not be feasible in real life.

## 1.1 Our Results

In this paper, we analyze property testing of joint distributions in the subcube conditioning model. Informally, the subcube conditioning model can be described in the following way. Let  $\Sigma^n$  be the domain of the distribution  $\mu$ . The Subcube Conditioning Oracle accepts  $A_1, A_2, \dots, A_n \subseteq \Sigma$  and constructs  $S = A_1 \times A_2 \times \dots \times A_n$  as the condition set. The oracle returns a vector  $x = (x_1, x_2, \dots, x_n)$ , where each  $x_i \in A_i$ , with probability  $\mu(x)/(\sum_{w \in S} \mu(w))$ . If  $\mu(S) = 0$ , we assume the oracle returns an element from  $S$  uniformly at random. We will call these kind of sample a subcube-conditional-sample and the corresponding sample complexity we will call subcube-conditional-sample complexity. There is no restriction on the individual  $A_i$ s. They may be unstructured or structured as pairs or intervals as used in [6, 7].

### Possible Application

We argue that the subcube conditional sampling model is natural and widely used in practice. We list some possible applications below.

**Side Channel Cryptanalysis.** In modern Cryptography, schemes are often “proven” secure under the assumption that the keys, internal randomness, and internal memory are inaccessible to the adversary. However, in practice Cryptographic schemes are deployed in wide variety of devices, specifically hand-held devices and smart cards. This situation leads to the “side channel attacks” where tampering with the keys or internal randomness is feasible. Specifically, the cryptanalytic techniques of fault attacks fixes/modifies some bits and test the resulting distributions. Our results show that constructing pseudorandom generators which can withstand *arbitrary* tampering with internal state, is impossible to achieve, implying the need of stronger assumptions for tamper-resilient cryptography.

**Verification of Random SAT solutions.** In software verification and related areas, random solutions to SAT problems are often used as a backbone. However testing whether the solution that one algorithm generates is actually uniform is a very important problem. Unfortunately the standard algorithms requires impractical complexity. Recently, Chakraborty et. al. [8] used the conditional sampling model to get a practically deployable solution. The model of subcube conditioning would be very effective to this problem as one natural conditioning technique is to fix some variables of the SAT equation and then test the distribution of the provided solution.

We remark that the idea of subcube conditioning has been mentioned in the literature. In fact, analysis of joint distributions using subcube conditioning was posed as an open problem in [6].

### Upper Bound Results

We focus on three fundamental properties of distributions: given a joint distribution  $\mu$  over  $\Sigma^n$  we would like to test, using subcube-conditional-samples, (a) if  $\mu$  is uniform, (b)  $\mu$  is identical to a known distribution, and (c) if  $\mu$  is a product distribution. We have the following three theorems:

**Theorem 1.1.** (Informal) *Let  $\mu$  is a probability distribution over  $\Sigma^n$ . There exists an algorithm for testing if  $\mu$  is uniform, using  $\tilde{O}(n^3 \epsilon^{-3})$  subcube-conditional-samples.*

**Theorem 1.2.** (Informal) Let  $\mu$  be a known probability distribution over the set  $\Sigma^n$ . Let  $\mu'$  be an unknown distribution over  $\Sigma^n$ . There exists an algorithm to test identity of  $\mu'$  with  $\mu$  using  $\tilde{O}(n^3\epsilon^{-3})$  subcube-conditional-samples.

**Theorem 1.3.** (Informal) Let  $\mu$  be a probability distribution over the set  $\Sigma^n$ . There exists an algorithm to test whether  $\mu$  is a product distribution using  $\tilde{O}(n^6 \log \log |\Sigma| \epsilon^{-5})$  subcube-conditional-samples.

Let us start with the problem of testing if a given distribution is uniform. This is in fact the central problem in testing various other properties of distribution. Let  $\mu$  be a distribution over  $\Sigma^n$  with marginals  $\mu_1, \dots, \mu_n$ .

The simplest case is when  $\mu$  is a product of  $n$  independent distributions. that is,  $\mu_i$ 's are independent but not necessarily identical. But if  $\mu$  is  $\epsilon$ -far from uniform one expects to find at least one  $\mu_i$  which is  $\epsilon/\text{poly}(n)$ -far from uniform. Then one can use the algorithm from [6, 7] to test if  $\mu_i$  using  $\text{poly}(n)$  subcube-conditional-samples.

But if the  $\mu_i$ 's are not independent then it is possible that all the individual marginals are uniform but still the  $\mu$  is  $\epsilon$ -far from uniform. For this case we define a notion of “conditional distance”. We show that there exists at least one  $i$  such that the expected “conditional distance” of  $\mu_i$  from uniform is more than  $\epsilon/\text{poly}(n)$ . We can again use the tester from [6, 7] to obtain a tester that can test if  $\mu$  is uniform using  $\text{poly}(n)$  subcube-conditional samples. The central idea for this case is the correct definition of the “conditional distance” and the “Chaining Lemma” that proves that such an  $i$  exists. Although the proof of “Chaining Lemma” (given in Section 3) is simple in hindsight, it is a powerful tool that acts as the central backbone for all our upper-bound proofs.

	Subcube-conditional Sample Complexity
Uniformity (Product of independent distribution)	$\tilde{O}(n^2\epsilon^{-2})$
Uniformity	$\tilde{O}(n^3\epsilon^{-3})$
Identity to known distribution	$\tilde{O}(n^3\epsilon^{-3})$
Independence Testing	$\tilde{O}(n^6 \log \log  \Sigma  \epsilon^{-5})$

Table 1: Summary of Our Upper Bound Results

## Lower Bound Results

We also present a lower bound for uniformity testing using the subcube conditioning oracle. Our lower bound holds even when the distribution  $\mu$  is a product of independent distributions. Note that since this case is the most basic case, this lower bound also holds for testing identity to a known distribution, or testing identity between two unknown distributions.

**Theorem 1.4.** (Informal) There is a product distribution  $\mu$  over the set  $\Sigma^n$  such that any algorithm to test whether  $\mu$  is uniform requires  $\Omega(\sqrt[4]{n})$  subcube-conditional samples.

The lower bound proof is more involved than the upper bound proofs. Although we use the usual techniques for proving lower bound, bounding the variation distance of product distributions turned out to be a tricky job. The main issue is that the relationship between the variation distance between two product distributions in terms of the distances of the individual marginal distributions is a less understood area. The main To get around this, we heavily use the Hellinger’s Distance. Our technique crucially depends on total variation distance and Hellinger distance to obtain upper and lower bounds on the variation distance between two product distributions in terms of the distances of the individual marginals. Some of the bounds (between Hellinger Distance and Variation Distance) proved for this lower bound can be of independent interest.

### 1.1.1 Open Problems

Although our lower and upper bounds are polynomially related to each other, we believe the lower bound can be improved to be almost linear in the dimension. However, it seems, to obtain an improved lower bound, one would require significantly new insight regarding the distance between product distributions in terms of their marginals. We leave it as an open problem for future work.

### Organization of the paper

The rest of the paper is organized in the following way. In Section 2, we define the notion of conditional distance and SubCube Conditioning. The elegant Chain Rule is described in Section 3. In Section 4 we present the identity tester and the derived uniformity tester followed by the special case of product distributions in Section 4.2. In Section 5, the tester for independence of marginals is described. Finally, we prove the lower bound in Section 6. The technical proofs for many of our lemmas and claims has been given in Section A and Section B.

## 2 Notations and Preliminaries

If  $S$  is a set  $|S|$  denotes the size of the set. If  $x$  is a vector of length  $n$ ,  $x_i$  denotes the  $i^{th}$  element of  $x$ .  $x^{(i)}$  denotes the substring of first  $i$  elements of  $x$ ;  $x^{(i)} = (x_1, x_2, \dots, x_i)$ . We denote the  $n$ -th harmonic number by  $H(n)$ .

For any set  $\Omega$  we denote by  $\mathcal{U}_\Omega$  the uniform distribution with support  $\Omega$ . In most of the cases the support of the distribution would be clear from the context and in that case we would drop the subscript and use  $\mathcal{U}$  as the uniform distribution over the support in question.

If  $\mu$  is a distribution with support  $\Omega$ , for any  $x \in \Omega$  we will denote by  $\Pr_\mu(x)$  the probability the  $x$  occurs when a random sample is drawn from  $\Omega$  according to  $\mu$ . If  $\mu$  is a joint distribution,  $\mu_i$  denotes the  $i^{th}$  marginal distribution of  $\mu$ .

If  $\mu$  is a distribution over  $\sigma^n$  with the marginals  $\mu_1, \dots, \mu_n$  and if the marginals are independent (that is,  $\mu$  is a product distribution) then we would write  $\mu = \mu_1 \otimes \dots \otimes \mu_n$ .

**TOTAL VARIATION DISTANCE.** Let  $\mu, \mu'$  be two distributions with support  $\Omega$ . The variation distance between  $\mu$  and  $\mu'$  denoted by  $d(\mu, \mu')$  is defined as

$$d(\mu, \mu') := \frac{1}{2} \sum_{x \in \Omega} \left| \Pr_\mu(x) - \Pr_{\mu'}(x) \right|.$$

We say  $\mu$  and  $\mu'$  are  $\epsilon$ -far (or  $\mu$  is  $\epsilon$ -far from  $\mu'$ ), when  $d(\mu, \mu') \geq \epsilon$ .

If  $\mu$  is a distribution with support  $\Omega$  and  $A \subseteq \Omega$ , then by  $(\mu \mid A)$  we denote the distribution over the support  $A$ . For any  $x \in A$  the probability that  $x$  occurs when a random sample is drawn from  $A$  (according to the distribution  $(\mu \mid A)$ ) is given by

$$\Pr_{\mu \mid A}(x) = \frac{\Pr_\mu(x)}{\sum_{y \in A} \Pr_\mu(y)}.$$

**HELLINGER DISTANCE.** Let  $\mu, \mu'$  be two distributions with support  $\Omega$ . The Hellinger distance between  $\mu$  and  $\mu'$  denoted by  $H(\mu, \mu')$  is defined as

$$H(\mu, \mu') = \frac{1}{\sqrt{2}} \sqrt{\sum_{x \in \Omega} \left( \sqrt{\Pr_\mu(x)} - \sqrt{\Pr_{\mu'}(x)} \right)^2} = \sqrt{\left( 1 - \sum_{x \in \Omega} \sqrt{\Pr_\mu(x) \Pr_{\mu'}(x)} \right)}$$

Hellinger distance has some nice properties and is useful for bounding lower and upper bounding variation distance.

$$d(\mu, \mu') \leq 2H(\mu, \mu') \leq 2\sqrt{d(\mu, \mu')}$$

Also for any two product distributions  $\mu = \mu_1 \otimes \cdots \otimes \mu_n$  and  $\mu' = \mu'_1 \otimes \cdots \otimes \mu'_n$

$$H(\mu, \mu')^2 \leq \sum_{i=1}^n H(\mu_i, \mu'_i)^2.$$

**CONDITIONAL DISTANCE.** Let  $\mu, \mu'$  be two distributions over  $\Omega$ . Let  $A \subseteq \Omega$ . The variation distance between  $\mu$  and  $\mu'$  conditioned on  $A$  (denote by  $d(\mu, \mu'|A)$ ) is defined as

$$d(\mu, \mu'|A) := \sum_{x \in \Omega} \left| \Pr_{\mu|A}(x) - \Pr_{\mu'|A}(x) \right|.$$

We say  $\mu$  and  $\mu'$  are  $\epsilon$ -far, conditioned on  $A$ , when  $d(\mu, \mu'|A) \geq \epsilon$ .

**SUBCUBE CONDITIONING.** In this paper we work with joint distributions;  $\Omega = \Sigma^n$  for some set  $\Sigma$ . We consider conditional distance under the condition on  $A = A_1 \times A_2 \times \cdots \times A_n$  where each  $A_i \subseteq \Sigma$ .

Let  $\mu$  be a distribution over  $\Sigma^n$  and  $X = (X_1, X_2, \dots, X_n)$  be a random variable distributed according to  $\mu$ .  $\mu^{(i)}$  denotes the distribution over  $\Sigma^i$  where for every  $x \in \Sigma^i$ ,

$$\Pr_{\mu^{(i)}}(x) = \Pr_{X \sim \mu}[(X_1, X_2, \dots, X_i) = (x_1, x_2, \dots, x_i)].$$

Let  $w \in \Sigma^j$  for some  $j < i$ .  $\mu_i | w$  denotes the marginal distribution  $\mu_i$  when first  $j$  random variables fixed to  $w$ .

$$\Pr_{\mu_i | w}(x) = \Pr_{X \sim \mu}[X_i = x | \bigwedge_{k=1}^j X_k = w_k].$$

Let  $\mu, \mu'$  be two distributions over  $\Sigma^n$ . The conditional marginal distance of  $\mu_i$  and  $\mu'_i$  conditioned on  $w$  is given by

$$d(\mu_i, \mu'_i | w) = \sum_{x \in \Sigma} \left| \Pr_{\mu_i | w}(x) - \Pr_{\mu'_i | w}(x) \right|$$

The average conditional distance between  $\mu_i$  and  $\mu'_i$  is defined by

$$\mathbb{E}_{w \sim \mu^{(i-1)}}[d(\mu_i, \mu'_i | w)] = \sum_{w \in \Sigma^{i-1}} \Pr_{\mu^{(i-1)}}(w) d(\mu_i, \mu'_i | w).$$

## The SubCube Condition Model

Let  $\mu$  be a distribution over  $\Sigma^n$ . A *subcube conditional oracle* for  $\mu$ , denoted  $\text{SUBCOND}_\mu$ , takes as input a sequence of sets  $\{A_i\}_{i \in [n]}$ ,  $A_i \subseteq \Sigma$ . Let  $A$  be the product set  $A_1 \times \cdots \times A_n$ . The oracle returns an element  $x \in \Sigma^n$  with probability  $\frac{\Pr_\mu[x]}{\sum_{x \in A} \Pr_\mu[x]}$  independently of all previous calls to the oracle.

An  $(\epsilon, \delta)\text{SUBCOND}$  tester for a property  $\mathcal{P}$  with conditional sample complexity  $t$  is a randomized algorithm, that receives  $\epsilon, \delta > 0$ ,  $n \in \mathbb{N}$  and oracle access to  $\text{SUBCOND}_\mu$ , and operates as follows.

1. In every iteration, the algorithm (possibly adaptively) generates a set  $A \subseteq \Sigma^n$ , based on the transcript and its internal coin tosses, and calls the conditional oracle with  $A$  to receive an element  $x$ , drawn according to the distribution  $\mu$  conditioned on  $A$ .

2. Based on the received elements and its internal coin tosses, the algorithm accepts or rejects the distribution  $\mu$ .
3. The algorithm makes at most  $t$  queries to  $\text{SUBCOND}_\mu$ , where  $t$  can depend on  $\epsilon, \delta, \Sigma$  and  $n$ .

If  $\mu$  satisfies  $\mathcal{P}$  then the algorithm must accept with probability at least  $1 - \delta$ , and if  $\mu$  is  $\epsilon$ -far from all distributions satisfying  $\mathcal{P}$ , then the algorithm must reject with probability at least  $1 - \delta$ .

We will call such a tester an  $(\epsilon, \delta)\text{SUBCOND } \mathcal{P}$ -tester. For example a  $(\epsilon, \delta)\text{SUBCOND}$  Uniformity-tester is a  $(\epsilon, \delta)\text{SUBCOND}$  tester that tests if the given distribution is uniform, a  $(\epsilon, \delta)\text{SUBCOND}$  Identity-tester is a  $(\epsilon, \delta)\text{SUBCOND}$  tester that tests if the given distribution is identical to a known distribution and a  $(\epsilon, \delta)\text{SUBCOND}$  Product-tester is a  $(\epsilon, \delta)\text{SUBCOND}$  tester that tests if the given distribution is a product distribution or far from all the product distributions.

### 3 Chain Rule of Conditional Distances

Let  $\mu$  and  $\mu'$  be two distributions over  $\Sigma^n$ , and let  $X = (X_1, X_2, \dots, X_n)$  and  $X' = (X'_1, X'_2, \dots, X'_n)$  be the corresponding random variables. For any  $1 \leq i \leq n$ , we denote by  $\mu_i$  and  $\mu'_i$  the distributions of the  $i$ th marginals of  $\mu$  and  $\mu'$  respectively.

**Lemma 3.1** (Chain Rule of Conditional Distances). *Let  $\mu$  and  $\mu'$  be two distributions over  $\Sigma^n$ , and let  $X = (X_1, X_2, \dots, X_n)$  and  $X' = (X'_1, X'_2, \dots, X'_n)$  be two random variables with distribution  $\mu$  and  $\mu'$  respectively. Then the following holds.*

$$d(\mu, \mu') \leq d(\mu_1, \mu'_1) + \sum_{i=2}^n \mathbb{E}_{w \sim \mu^{(i-1)}} [d(\mu_i, \mu'_i | w)]$$

*Proof of Lemma 3.1:* <sup>1</sup> Let  $w = (w_1, w_2, \dots, w_n) \in \Sigma^n$ .

Let  $2 \leq i \leq n$ . Recall that  $w^{(i)}$  denotes the substring of first  $i$  elements of  $w$ .

$$\begin{aligned} d(\mu^{(i)}, \mu'^{(i)}) &= \sum_{w \in \Sigma^i} \left| \Pr_{\mu^{(i)}}(w) - \Pr_{\mu'^{(i)}}(w) \right| \\ &= \sum_{w \in \Sigma^i} \left| \Pr_{X \sim \mu} [\wedge_{j=1}^{i-1} X_j = w_j] \Pr_{X \sim \mu} [X_i = w_i | \wedge_{j=1}^{i-1} X_j = w_j] \right. \\ &\quad \left. - \Pr_{X' \sim \mu'} [\wedge_{j=1}^{i-1} X'_j = w_j] \Pr_{X' \sim \mu'} [X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j] \right| \\ &\leq \sum_{w \in \Sigma^i} \left| \Pr_{X \sim \mu} [\wedge_{j=1}^{i-1} X_j = w_j] \left( \Pr_{X \sim \mu} [X_i = w_i | \wedge_{j=1}^{i-1} X_j = w_j] - \Pr_{X' \sim \mu'} [X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j] \right) \right| + \\ &\quad \sum_{w \in \Sigma^i} \left| \Pr_{X' \sim \mu'} [X'_i = w_i | \wedge_{j=1}^{i-1} X'_j = w_j] \left( \Pr_{X \sim \mu} [\wedge_{j=1}^{i-1} X_j = w_j] - \Pr_{X' \sim \mu'} [\wedge_{j=1}^{i-1} X'_j = w_j] \right) \right| \end{aligned}$$

---

<sup>1</sup>While one may consider the coupling technique to be a natural tool to prove the chaining lemma, intricacies regarding conditional coupling is not completely clear to us.

Now, the second term reduces to,

$$\begin{aligned}
& \sum_{w \in \Sigma^i} \left| \Pr[X'_i = w_i \mid \wedge_{j=1}^{i-1} X'_j = w_j] \left( \Pr_{X \sim \mu} [\wedge_{j=1}^{i-1} X_j = w_j] - \Pr_{X' \sim \mu'} [\wedge_{j=1}^{i-1} X'_j = w_j] \right) \right| \\
&= \sum_{w' \in \Sigma^{i-1}} \left| \Pr_{X \sim \mu} [\wedge_{j=1}^{i-1} X_j = w'_j] - \Pr_{X' \sim \mu'} [\wedge_{j=1}^{i-1} X'_j = w'_j] \right| \sum_{w_i \in \Sigma} \Pr_{X' \sim \mu'} [X'_i = w_i \mid \wedge_{j=1}^{i-1} X'_j = w'_j] \\
&= \sum_{w' \in \Sigma^{i-1}} \left| \Pr_{X \sim \mu} [\wedge_{j=1}^{i-1} X_j = w'_j] - \Pr_{X' \sim \mu'} [\wedge_{j=1}^{i-1} X'_j = w'_j] \right| \\
&= \sum_{w' \in \Sigma^{i-1}} \left| \Pr_{\mu^{(i-1)}}(w') - \Pr_{\mu'^{(i-1)}}(w') \right|
\end{aligned}$$

The second equality follows from the fact that for each  $w' \in \Sigma^{i-1}$ ,  $\sum_{w_i \in \Sigma} \Pr[X'_i = w_i \mid \wedge_{j=1}^{i-1} X'_j = w'_j] = 1$ .<sup>2</sup> Hence,

$$d(\mu^{(i)}, \mu'^{(i)}) \leq d(\mu^{(i-1)}, \mu'^{(i-1)}) + \sum_{w \in \Sigma^{i-1}} \Pr_{\mu^{(i-1)}}(w) d(\mu_i, \mu'_i | w)$$

Solving the recursion we get the lemma.  $\square$

Arranging the marginals by the increasing order of the average conditional distance, we get the immediate corollary.

**Lemma 3.2.** *If  $d(\mu, \mu') \geq \epsilon$ , then there exists a  $c \in \mathbb{N}$  such that*

$$2^{c-1} \leq \left| \left\{ i \in [n] \mid \mathbb{E}_{w \sim \mu^{(i-1)}} [d(\mu_i, \mu'_i | w)] \geq \frac{\epsilon}{2^c H(n)} \right\} \right|$$

We refer the reader to Appendix A.1 for the proof.

## 4 Testing Identity with a known distribution

In this section, we present an identity tester of Sample complexity  $\tilde{O}(n^3 \epsilon^{-3})$ . We recall the following result proved in [9]

**Lemma 4.1.** [9] *Let  $\mu$  be a known distribution over  $\Sigma$ . Given  $0 < \epsilon < 1$  and  $0 < \delta < 1$  and a distribution  $\mu'$  over  $\Sigma$  there is an adaptive  $(\epsilon, \delta)$ -Identity Tester with conditional sample complexity  $\tilde{O}(\frac{1}{\epsilon^2} \log(\frac{1}{\delta}))$ . In other words, there is a tester that draws  $\tilde{O}(\frac{1}{\epsilon^2} \log(\frac{1}{\delta}))$  number of conditional samples and*

- if  $\mu = \mu'$ , then the tester will accept with probability  $(1 - \delta)$ , and
- if  $d(\mu, \mu') \geq \epsilon$  then the tester will reject with probability  $(1 - \delta)$ .

Let  $\mu$  be a known distribution over  $\Sigma^n$ ,  $\mu'$  be an unknown distribution over  $\Sigma^n$  that can be accessed via  $\text{SUBCOND}_{\mu'}$  oracle, and  $\epsilon$  be the target distance. The following algorithm tests identity of  $\mu'$  with  $\mu$ . We use the identity tester **BasicIDTester** over  $\Sigma$  guaranteed by Lemma 4.1 as a subroutine. Fix  $\delta = 1/3$ .

---

<sup>2</sup>If  $w'$  is outside of support of  $\mu'$ , like in [7], we can define the conditional probability to be uniform over  $\Sigma$



---

**Algorithm 1** The Identity Tester for Joint Distributions

---

```
1:  $\epsilon' = \epsilon/2^{j+1}H(n)$ 
2:  $\delta' = \delta\epsilon/24n(\log n)^2$ 
3: for  $j = 1$  to  $\log n + 1$  do
4:   Create a set  $S_j$  by sampling, with replacement,  $(4n/2^j)$  element from  $[n]$  uniformly at random.
5:   for all  $i \in S_j$  do
6:     for  $k' = 1$  to  $3/\epsilon'$  do
7:       Sample  $w \sim \mu$ . Let  $w = (w_1, \dots, w_n)$ .
8:       Consider the distribution  $\mu_i | w^{(i-1)}$ .
9:       If BasicIDTester $(\mu_i | w^{(i-1)}, \mu'_i | w^{(i-1)}, \epsilon', \delta')$  rejects, Output REJECT
10:    end for
11:  end for
12: end for
13: Output ACCEPT
```

---

**Theorem 4.2.** Let  $\delta = \frac{1}{3}$ . Algorithm 1 is a  $(\epsilon, \delta)$  SUBCOND Identity Tester for joint distributions with conditional sample complexity of  $\tilde{O}(n^3/\epsilon^3)$  where  $\tilde{O}$  hides a polynomial function of  $\log n, \log \epsilon$ .

*Proof.* The  $(\epsilon', \delta')$  **BasicIDTester** needs conditional samples for testing whether  $d(\mu_i, \mu'_i | w^{(i-1)}) \geq \epsilon/2^{j+1}H(n)$ . To answer the conditional queries with condition  $B \subseteq \Sigma$  for the distribution  $\mu'_i | w^{(i-1)}$ , we set  $A_j = \{w_j\}$  for  $j = 1, 2, \dots, i-1$ ,  $A_i = B$ , and  $A_j = \Sigma$  for  $j = i+1, \dots, n$ , and query the SUBCOND oracle with the condition  $A$ . This correctly simulates the conditional oracle required by the underlying identity tester. Thus Algorithm 1 is a SUBCOND Tester.

Next, we prove the sample complexity of Algorithm 1 as claimed in the Theorem 4.2.

**Sample Complexity of Algorithm 1:** By Lemma 4.1, a query to **BasicIDTester** $(\mu_i | w^{(i-1)}, \mu'_i | w^{(i-1)}, \epsilon', \delta')$  requires  $\tilde{O}(1/\epsilon'^2) = \tilde{O}(2^{2j+2}H(n)^2/\epsilon^2)$  samples. For each index in  $S_j$ , the basic Identity tester is queried  $2^{j+1}3H(n)/\epsilon$  times. Hence, the total sample complexity of our tester is

$$\sum_{j=1}^{\log n+1} \frac{4n}{2^j} \times \frac{2^{j+1}3H(n)}{\epsilon} \times \tilde{O}\left(\frac{2^{2j+2}H(n)^2}{\epsilon^2}\right) = \tilde{O}\left(\frac{nH(n)^3}{\epsilon^3}\right) \sum_{j=1}^{\log n+1} 2^{2j} = \tilde{O}(n^3/\epsilon^3)$$

The correctness of Algorithm 1 is proved in Appendix A.2. □

## 4.1 Uniformity Tester for Arbitrary Joint Distribution

If we set  $\mu$  to be the uniform distribution, then Algorithm 1 gives us a Uniformity Tester. Hence, we get the following as a corollary of Theorem 4.2.

**Theorem 4.3.** Let  $\delta = \frac{1}{3}$ . There exists a  $(\epsilon, \delta)$ - Uniformity Tester for any joint distribution with conditional sample complexity of  $\tilde{O}(n^3/\epsilon^3)$  where  $\tilde{O}$  hides a polynomial function of  $\log n, \epsilon$ .

## 4.2 Uniformity Testing for Product Distributions

If  $\mu$  is a product distribution, then we can achieve better sample complexity. In a product distribution the marginals are independent. Thus Lemma 3.1 gives us

$$d(\mu, \mu') \leq \sum_{i=1}^n d(\mu_i, \mu'_i)$$

Similar to Algorithm 1, we sample marginals and check the uniformity of individual marginals using the uniformity tester over  $\Sigma$ , designed in [7] and [6]. However, in this case we do not need to sample  $w$  and condition the marginals. This gives us an uniformity tester with  $\tilde{O}(n^2\epsilon^{-2})$  sample complexity.

**Theorem 4.4.** *Let  $\delta = 1/3$ . If  $\mu$  is a distribution over  $\Sigma^n$  and the  $\mu_i$ s are independent, then there exists an  $(\epsilon, \delta)$  SUBCOND Uniformity Tester with conditional sample complexity of  $\tilde{O}(n^2/\epsilon^2)$ , where the  $\tilde{O}$  hides a polynomial function of  $\log n$  and  $\epsilon$ .*

For the algorithm and other details, we refer the reader to Appendix A.3.

## 5 Testing Independence of Marginals

Let  $\mu$  be a probability distribution over  $\Sigma^n$ . In this section we present an algorithm to test whether  $\mu$  is a product distribution; i.e. *whether all the marginals of  $\mu$  are independent or  $\mu$  is far from all the product distributions.*

Define  $\mu'$  to be the product of marginals of  $\mu$ .

$$\Pr_{\mu'}(w) = \prod_{i=1}^n \Pr_{\mu_i}(w_i) \quad \forall w \in \Sigma^n$$

By definition, the marginal distributions  $\mu'_i$  are exactly the marginal distribution  $\mu_i$ . If  $\mu$  is  $\epsilon$ -far from all the product distribution, it is  $\epsilon$ -far from  $\mu'$ . Using the Chain Rule (Lemma 3.1),

$$\begin{aligned} d(\mu, \mu') &\leq d(\mu_1, \mu'_1) + \sum_{i=2}^n \mathbb{E}_{w \sim \mu^{(i-1)}} [d(\mu_i, \mu'_i | w)] \\ &= \sum_{i=2}^n \sum_{w \in \Sigma^{i-1}} \Pr_{\mu^{(i-1)}}(w) \left( \sum_{w_i \in \Sigma} \left| \Pr_{\mu_i}(w_i | w) - \Pr_{\mu'_i}(w_i) \right| \right) \\ &= \sum_{i=2}^n \sum_{w \in \Sigma^{i-1}} \Pr_{\mu^{(i-1)}}(w) d(\mu_i | w, \mu_i) \end{aligned}$$

Therefore, We need to test whether there exists  $i \in [n]$ , such that the marginal distribution  $\mu_i$  is far (on average) from the conditional marginal distribution  $\mu_i | w$ . Unfortunately, in this case, both the distributions are unknown. Luckily, as both  $\mu_i$  and  $\mu_i | w$  is distributed over  $\Sigma$ , we can use unknown distribution tester from [9], where identity between two unknown distribution is tested using  $\tilde{O}(\log \log |\Sigma|/\epsilon^5)$  sample complexity. The only thing left is to sample  $w$  according to  $\mu^{i-1}$ . The algorithm queries the oracle with  $\Sigma^n$  as the condition, and selects the first  $i - 1$  bit as the sample  $w$ .

**Theorem 5.1.** *Let  $\delta = \frac{1}{3}$ . There exists an  $(\epsilon, \delta)$  SUBCOND Product Tester for joint distributions with conditional sample complexity of  $\tilde{O}(n^6 \log \log |\Sigma| \epsilon^{-5})$  where  $\tilde{O}$  hides a polynomial function of  $\log n$ .*

The algorithm and the proof details are presented in Appendix A.4.

## 6 Lower Bounds

**Theorem 6.1.** *For any  $0 \leq \epsilon \leq 1$  any  $(\epsilon, 1/3)$  – SUBCOND Uniformity-Tester has subcube-conditional sample complexity  $\Omega(\sqrt[4]{n})$ . The lower bound holds even for the case when the domain is  $\{0, 1\}^n$  and the given distribution is a product of  $n$  independent (though not necessarily identical) distributions.*

*Proof.* Let  $\mu$  be a product distributions over the domain  $\{0, 1\}^n$  with marginals  $\mu_1, \dots, \mu_n$ . So  $\mu = \mu_1 \otimes \dots \otimes \mu_n$ . Note that since the  $\mu_i$  are independent, if  $i \neq j$  then conditioning on  $\mu_i$  does not affect the samples we get from a  $\mu_j$ . Also since the  $\mu_i$  are all distributions over a two element set (namely  $\{0, 1\}$ ) conditioning on any subset of  $\{0, 1\}$  also of no use. Thus drawing subcube-conditional-samples from  $\mu$  is as good as drawn samples (without any conditioning) from  $\mu$ .

So it is sufficient for us to prove that for any  $0 \leq \epsilon \leq 1$  any  $(\epsilon, 1/3)$  Uniformity-Tester has sample complexity  $\Omega(\sqrt[4]{n})$ , when the domain is  $\{0, 1\}^n$  and the given distributions are product distributions.

The main idea of the proof is to use a standard technique from property testing where the following lemma is used. The following lemma has been rewritten in the language and context of this paper. A proof of the general statement of the lemma can be found in [10, 11].

**Theorem 6.2.** *Let  $P$  be a property of distributions over  $\sigma^n$  that we want to test. Suppose  $\mathcal{D}_Y$  is a distribution over all the distribution that satisfy the given property  $P$ , and let  $\mathcal{D}_N$  be a distribution over all distributions that are  $\epsilon$ -far from satisfying the property  $P$ . Let  $Q_Y$  be the distribution over outcomes of  $q$  samples when the samples are drawn from a distribution  $\mathcal{D}_Y$ , that is drawn according according to  $\mathcal{D}_Y$ . Similarly, let  $Q_N$  be the distribution over outcomes of  $q$  samples when the samples are drawn from a distribution  $\mathcal{D}_N$ , that is drawn according to the  $\mathcal{D}_N$ . If the variation distance between  $Q_Y$  and  $Q_N$  is less than  $1/3$  then any  $(\epsilon, 1/3)$ -Tester for the property  $P$  will have sample complexity more than  $q$ .*

In the context of our theorem we have the property  $P$  is “Uniformity”. So the distribution  $\mathcal{D}_Y$  is the uniform distribution over the domain  $\{0, 1\}$ . Now let us define the distribution  $\mathcal{D}_N$ :

Let  $D_1$  be the distribution over  $\{0, 1\}$  where 1 is produced with probability  $(1/2 + 2\sqrt{\frac{\epsilon}{n}})$  and 0 produced with probability  $(1/2 - 2\sqrt{\frac{\epsilon}{n}})$ . And let  $D_0$  be the distribution over  $\{0, 1\}$  where 1 is produced with probability  $(1/2 - 2\sqrt{\frac{\epsilon}{n}})$  and 0 produced with probability  $(1/2 + 2\sqrt{\frac{\epsilon}{n}})$ .

Consider the set of distributions  $\mathcal{D}$  over  $\{0, 1\}^n$  which are a product of  $n$  distribution each of which is either  $D_0$  or  $D_1$ . That is,

$$\mathcal{D} = \{\mu_1 \otimes \dots \otimes \mu_n \mid \text{for all } i, \mu_i \text{ is either } D_0 \text{ or } D_1\}$$

**Claim 6.3.** *Any  $\mu \in \mathcal{D}$  is  $\epsilon$ -far from uniform. That is, for any  $\mu \in \mathcal{D}$  we have*

$$d(\mu, \mathcal{U}) \geq \epsilon$$

From the Claim 6.3 we see that all the distributions in  $\mathcal{D}$  are  $\epsilon$ -far from uniform. Thus we can take the distribution  $\mathcal{D}$  as our distribution  $\mathcal{D}_N$ . If a distribution is drawn from  $\mathcal{D}_N$  or  $\mathcal{D}_Y$ ,  $q$  samples from the distribution would give  $q$  many  $\{0, 1\}$ -strings of length  $n$ . Note that if a distribution is drawn from  $\mathcal{D}_Y$  (that is, the distribution is the uniform distribution over  $\{0, 1\}^n$ ) then the distribution of the outcomes of  $q$  samples is a uniform distribution over  $\{0, 1\}^{nq}$ . So, by theorem 6.2, it is enough to show that if  $\mu$  is drawn from  $\mathcal{D}_N$  then the distribution of the outcomes (as a distribution over  $\{0, 1\}^{nq}$ ) is  $1/3$ -close to uniform.

Note that  $\mu$  is a distribution drawn from  $\mathcal{D}_N$  we can think of  $\mu$  as  $\mu_1 \otimes \dots \otimes \mu_n$  where each  $\mu_i$  is independently and uniformly choose from the set  $\{D_0, D_1\}$ . Let  $\mu^q$  be the distribution over  $\{0, 1\}^{nq}$  when  $q$  samples are drawn from  $\mu$ . And now the following lemma completes the proof of Theorem 6.1.

**Lemma 6.4.** *If  $q \leq \sqrt[4]{n}/20\sqrt{\epsilon}$  then*

$$d(\mu^q, \mathcal{U}) \leq \frac{1}{3}.$$

The proof of Claim 6.3 and Lemma 6.4 is given in Appendix B. □

## 7 Conclusion

In this paper we analyzed property testing of joint distributions in the conditional sampling model. We considered the natural subcube conditioning and presented testers to test uniformity, identity with a known distribution, and independence of marginals of query complexity polynomial in the dimension, thus avoiding the curse of dimensionality. We also presented a lower bound for uniformity testing in the subcube conditioning model.

## References

- [1] Jayadev Acharya, Constantinos Daskalakis, and Gautam Kamath. Optimal testing for properties of distributions. *CoRR*, abs/1507.05952, 2015.
- [2] Tugkan Batu, Sanjoy Dasgupta, Ravi Kumar, and Ronitt Rubinfeld. The complexity of approximating the entropy. *SIAM J. Comput.*, 35(1):132–150, 2005.
- [3] Tugkan Batu, Lance Fortnow, Eldar Fischer, Ravi Kumar, Ronitt Rubinfeld, and Patrick White. Testing random variables for independence and identity. In Bob Werner, editor, *Proceedings of the 42nd Annual Symposium on Foundations of Computer Science (FOCS-01)*, pages 442–451, Los Alamitos, CA, October 14–17 2001.
- [4] Tugkan Batu, Lance Fortnow, Ronitt Rubinfeld, Warren D. Smith, and Patrick White. Testing closeness of discrete distributions. *Journal of the ACM*, 60(1):4:1–4:25, February 2013.
- [5] Clément L. Canonne. A survey on distribution testing: Your data is big, but is it blue? *Electronic Colloquium on Computational Complexity (ECCC)*, 22:63, 2015.
- [6] Clement L. Canonne, Dana Ron, and Rocco A. Servedio. Testing probability distributions using conditional samples. *SIAM Journal on Computing*, 44(3):540–616, 2015.
- [7] Sourav Chakraborty, Eldar Fischer, Yonatan Goldhirsh, and Arie Matsliah. On the power of conditional samples in distribution testing. In *Innovations in Theoretical Computer Science, ITCS '13, Berkeley, CA, USA, January 9-12, 2013*, pages 561–580, 2013.
- [8] Sourav Chakraborty and Kuldeep Meel. Testing correctness of programs that claim to produce satisfying assignments uniformly at random. *Under Preparation*, 2016.
- [9] Moein Falahatgar, Ashkan Jafarpour, Alon Orlitsky, Venkatadheeraj Pichapati, and Ananda Theertha Suresh. Faster algorithms for testing under conditional sampling. In *Proceedings of The 28th Conference on Learning Theory, COLT 2015, Paris, France, July 3-6, 2015*, pages 607–636, 2015.
- [10] Eldar Fischer. The difficulty of testing for isomorphism against a graph that is given in advance. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing, Chicago, IL, USA, June 13-16, 2004*, pages 391–397, 2004.
- [11] Eldar Fischer, Ilan Newman, and Jirí Sgall. Functions that have read-twice constant width branching programs are not necessarily testable. *Random Struct. Algorithms*, 24(2):175–193, 2004.
- [12] Oded Goldreich and Dana Ron. On testing expansion in bounded-degree graphs. In Oded Goldreich, editor, *Studies in Complexity and Cryptography*, volume 6650 of *Lecture Notes in Computer Science*, pages 68–75. Springer, 2011.

- [13] Reut Levi, Dana Ron, and Ronitt Rubinfeld. Testing properties of collections of distributions. *Theory of Computing*, 9:295–347, 2013.
- [14] Siu on Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In Chandra Chekuri, editor, *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*, pages 1193–1203. SIAM, 2014.
- [15] Sofya Raskhodnikova, Dana Ron, Amir Shpilka, and Adam Smith. Strong lower bounds for approximating distribution support size and the distinct elements problem. *SIAM J. Comput.*, 39(3):813–842, 2009.
- [16] Gregory Valiant and Paul Valiant. A clt and tight lower bounds for estimating entropy. *Electronic Colloquium on Computational Complexity (ECCC)*, 2010.
- [17] Paul Valiant. Testing symmetric properties of distributions. *SIAM J. Comput.*, 40(6):1927–1968, 2011.

## A Leftout Proofs of the Upper Bounds

### A.1 Proof of Lemma 3.2

*Proof of Lemma 3.2.* Without loss of generality let  $i_1, i_2, \dots, i_n$  be indices such that

$$\mathbb{E}_{w \sim \mu^{(i_1-1)}}[d(\mu_{i_1}, \mu'_{i_1} | w)] \geq \mathbb{E}_{w \sim \mu^{(i_2-1)}}[d(\mu_{i_2}, \mu'_{i_2} | w)] \geq \mathbb{E}_{w \sim \mu^{(i_n-1)}}[d(\mu_{i_n}, \mu'_{i_n} | w)]$$

We will need the following Claim.

**Claim A.1.** *There exists  $k \in [n]$  such that*

$$\mathbb{E}_{w \sim \mu^{(i_k-1)}}[d(\mu_{i_k}, \mu'_{i_k} | w)] \geq \epsilon / kH(n)$$

*Proof of Claim A.1.* If no such  $k$  exists, then

$$d(\mu, \mu') \leq \sum_{k=1}^n \mathbb{E}_{w \sim \mu^{(i_k-1)}}[d(\mu_{i_k}, \mu'_{i_k} | w)] < \sum_{k=1}^n \epsilon / (kH(n)) \leq \epsilon$$

which contradicts the distance assumption in Lemma 3.2. □

Let  $k$  be the index from Claim A.1. We put  $c = \lceil \log k \rceil$  to get  $\epsilon / 2^c H(n) \leq \epsilon / kH(n)$ . Clearly

$$\left| \left\{ i \in [n] \mid \mathbb{E}_{w \sim \mu^{(i-1)}}[d(\mu_i, \mu'_i | w)] \geq \frac{\epsilon}{2^c H(n)} \right\} \right| \geq k \geq 2^{c-1}.$$

□

### A.2 Correctness of the Algorithm 1

Let Algorithm 1 queries the underlying tester  $q$  times.

$$q = \sum_{j=1}^{\log n + 1} \frac{4n}{2^j} \times \frac{2^{j+1} 3H(n)}{\epsilon} = 24nH(n) \sum_{j=1}^{\log n + 1} \frac{1}{\epsilon} \leq 24n (\log n)^2 / \epsilon$$

COMPLETENESS. We start by proving the completeness of the algorithm. We show, if  $\mu'$  is identical with  $\mu$  then with probability at least  $(1 - \delta)$  the algorithm ACCEPTS. We will show that if  $d(\mu, \mu') = 0$  is the algorithm will reject with probability at most  $\delta$ . Fix  $\delta = 1/3$ .

Algorithm 1, rejects  $\mu'$  if there exists  $i \in [n]$  and a sampled  $w = (w_1, \dots, w_n) \in \Sigma^n$  the underlying Identity Tester rejects in the Step 4.

Suppose  $\mu$  and  $\mu'$  are identical. Then for all  $w \in \Sigma^{i-1}$ ,  $\mu_i|w$  is identical to  $\mu'_i|w$ . For each query, BasicIDTester will reject in Step 9 with probability at most  $\delta\epsilon/24n(\log n)^2$ . Since the BasicIDTester is called at most  $24n(\log n)^2/\epsilon$  times, by union bound, if  $d(\mu, \mu') = 0$ , the algorithm will reject  $\mu'$  with at most  $\delta$  probability.

SOUNDNESS. Now we prove the soundness of the Algorithm 1. Let  $\mu$  be a distribution over  $\Sigma^n$  and  $d(\mu, \mu') \geq \epsilon$ . We shall show that Algorithm 1, rejects  $\mu'$  with probability at least  $2/3$ .

Let

$$\tau_c \stackrel{\text{def}}{=} \left\{ i \in [n] \mid \mathbb{E}_{w \sim \mu^{(i-1)}}[d(\mu_i, \mu'_i|w)] \geq \frac{\epsilon}{2^c H(n)} \right\}$$

By Lemma 3.2, there exist a  $c \leq \lceil \log n \rceil$ , such that  $|\tau_c| \geq 2^{c-1}$ .

The following lemma is a direct consequence of Lemma 3.2.

**Lemma A.2.** *Let  $\mu$  be a distribution over  $\Sigma^n$ , and  $\mu$  is  $\epsilon$ -far from uniform. Let  $X = (X_1, \dots, X_n)$  be a random variable with distribution  $\mu$ . Let  $w = (w_1, w_2, \dots, w_n)$  be a random sample drawn from  $\Sigma^n$  according to the distribution  $\mu$ .*

*Then for all  $i \in \tau_c$ ,*

$$\Pr_{w \sim \mu} \left[ d(\mu_i, \mu'_i \mid w^{i-1}) \geq \frac{\epsilon}{2^{c+1} H(n)} \right] \geq \frac{\epsilon}{2^{c+1} H(n)}$$

For each  $i \in \tau_c$ , define

$$\Gamma_i \stackrel{\text{def}}{=} \left\{ w \in \Sigma^{i-1} \mid d(\mu_i, \mu'_i \mid \bigwedge_{j=1}^{i-1} X_j = w_j) < \frac{\epsilon}{2^{c+1} H(n)} \right\}$$

Let  $S_j$  be the set of indices sampled in the Step 3 in the  $j^{\text{th}}$  iteration.

If Algorithm 1 fails to reject  $\mu'$ , one of the following three cases happen.

1. No index from  $\tau_c$  was sampled in  $S_j$ . Specifically,  $S_c \cap \tau_c = \emptyset$ . Probability of this event is

$$\left( 1 - \frac{|\tau|}{n} \right)^{\frac{4n}{2^c}} \leq e^{-2}.$$

2. For all index  $i \in S_c \cap \tau_c$ , all the sampled  $w$ s are from the set  $\Gamma_i$ . Probability of this event is

$$\left( 1 - \frac{\epsilon}{2^{c+1} H(n)} \right)^{\frac{2^{c+1} 3 H(n)}{\epsilon}} \leq e^{-3}.$$

3. For all index  $i \in S_c \cap \tau_c$ , for all the sampled  $w \notin \Gamma_i$ , underlying identity tester fails to reject. Probability of such event is at most  $\delta\epsilon/24n(\log n)^2$ , which is less than  $1/100$  for  $n \geq 2$ .

Hence, the probability that Algorithm 1 fails to reject  $\mu'$  is at most  $e^{-2} + e^{-3} + 1/100 < 1/3$ .

This completes the proof of Theorem 4.2.  $\square$

*Proof of Lemma A.2.* From Lemma 3.2, for all index  $i \in \tau_c$

$$\mathbb{E}_{w \sim \mu^{(i-1)}}[d(\mu_i, \mu'_i | w)] = \sum_{w \in \Sigma^{i-1}} \frac{\Pr(w)}{\mu^{i-1}} d(\mu_i, \mu'_i | w) \geq \frac{\epsilon}{2^c H(n)}$$

By simple averaging argument we conclude, if we sample according to  $\mu$ , with probability at least  $\frac{\epsilon}{2^{c+1}H(n)}$  we get a  $w \in \Sigma^{i-1}$ , such that

$$d(\mu_i, \mu'_i | w) \geq \frac{\epsilon}{2^{c+1}H(n)}.$$

□

### A.3 Proof Of Theorem 4.4

To obtain our “uniformity testers for product distributions” we need the following uniformity tester for the unstructured domains, proven in [7] and [6].

**Lemma A.3.** [7, 6] *Given  $0 < \epsilon < 1$  and  $0 < \delta < 1$  and a distribution  $\mu$  over  $\Sigma$  there is an adaptive  $(\epsilon, \delta)$ -Uniformity Tester with conditional sample complexity  $\tilde{O}(\frac{1}{\epsilon^2} \log(\frac{1}{\delta}))$ . In other words, there is a tester that draws  $\tilde{O}(\frac{1}{\epsilon^2} \log(\frac{1}{\delta}))$  number of independent conditional samples and*

- *if  $\mu$  is uniform over  $\Sigma$  then the tester will accept with probability  $(1 - \delta)$ , and*
- *if  $d(\mu, \mathcal{U}) \geq \epsilon$  then the tester will reject with probability  $(1 - \delta)$ .*

Algorithm 2 describes the uniformity tester when  $\mu_i$ s are independent. We use the conditional uniformity tester **BasicUniTester** guaranteed by Lemma A.3 as a subroutine. Fix  $\delta = 1/3$ .

---

**Algorithm 2** The Uniformity Tester for product of independent distributions

---

```

1:  $\epsilon' = \epsilon/2^j H(n)$ 
2:  $\delta' = \delta/8n$ 
3: for  $j = 1$  to  $\log n + 1$  do
4:   Create a set  $S_j$  by sampling, with replacement,  $(4n/2^j)$  element from  $[n]$  uniformly at random.
5:   for all  $i \in S_j$  do
6:     If BasicUniTester $(\mu_i, \mathcal{U}, \epsilon', \delta')$  rejects, Output REJECT
7:   end for
8: end for
9: Output ACCEPT
```

---

The **BasicUniTester** needs conditional samples (over  $\Sigma$ ) for testing if  $d(\mu_i, \mathcal{U}) \geq \epsilon/2^j H(n)$  in Step 4. Such a sample, conditioned on the set  $B$ , is obtained by drawing a sample from  $\mu$  conditioned on set  $A = (A_1 \otimes \cdots \otimes A_n) \subseteq \Sigma^n$  where  $A_i = B$  and  $A_j = \Sigma$  for all  $j \neq i \in [n]$ . Thus Algorithm 2 is a SUBCOND Uniformity Tester.

To prove Theorem 4.4 we need to show the correctness of the algorithm and the sample complexity as claimed.

### Sample Complexity of our Tester

From Lemma A.3 we have an  $(\epsilon/2^j H(n), \delta/8n)$ -Uniformity Tester that has sample complexity  $\tilde{O}(2^{2j} H(n)^2 \log(n\delta^{-1})\epsilon^{-2})$

Now in Algorithm 2 each  $S_j$  is of size  $4n/2^j$ , and for each element  $i$  in  $S_j$  the  $(\epsilon/2^j H(n), \delta/8n)$ -Uniformity Tester is used. So the sample complexity of the algorithm is

$$\sum_{j=1}^{\log n+1} \frac{4n}{2^j} \tilde{O}(2^{2j} H(n)^2 \log(n\delta^{-1})\epsilon^{-2}) = \tilde{O}(n^2/\epsilon^2).$$

### Proof of Correctness of the algorithm

Let us first prove the completeness. That is, if  $\mu$  is uniform over  $\Sigma^n$  then with probability at least  $(1 - \delta)$  the algorithm outputs ACCEPT. We will show that if  $\mu$  is uniform the algorithm will reject with probability at most  $\delta$ . Note that the algorithm ACCEPTS if for all  $i$  and  $j$  the  $(\epsilon/2^j H(n), \delta/8n)$ -Uniformity Tester does not reject in the Step 4. The important thing to note is that if  $\mu$  is uniform then  $\mu_i$  is also uniform. Hence for a given  $i$  and  $j$  the  $(\epsilon/2^j H(n), \delta/8n)$ -Uniformity Tester will reject in Step 4 with probability less than  $\delta/8n$ . Since the  $(\epsilon/2^j H(n), \delta/8n)$ -Uniformity Tester is called at most  $8n$  times, by union bound, if  $\mu$  is uniform the algorithm will reject  $\mu$  with at most  $\delta$  probability.

Now, we prove the soundness. We need to show that if  $d(\mu, \mathcal{U}) \geq \epsilon$ , then the Uniformity Tester of Algorithm 2 rejects with probability  $(1 - \delta) = 2/3$ .

Let

$$\tau_c = \left\{ i \in [n] \mid d(\mu_i, \mathcal{U}) \geq \frac{\epsilon}{2^c H(n)} \right\}.$$

As the marginals of  $\mu$  are independent, for all  $i \in [n]$ ,

$$\begin{aligned} \mathbb{E}_{w \sim \mu^{(i-1)}}[d(\mu_i, \mathcal{U}|w)] &= \sum_{w \in \Sigma^{i-1}} \Pr_{\mu^{i-1}}(w) d(\mu_i, \mathcal{U}|w) \\ &= \sum_{w \in \Sigma^{i-1}} \Pr_{\mu^{i-1}}(w) d(\mu_i, \mathcal{U}) \\ &= d(\mu_i, \mathcal{U}) \end{aligned}$$

Hence, by Lemma 3.2, there exists  $c \leq \lceil \log n \rceil$ , such that  $|\tau_c| \geq 2^{c-1}$ . We identify the above  $\tau_c$  as  $\tau$ .

Let  $S_c$  be the set of indices sampled in Step 2, when  $j = c$ .

The algorithm will REJECT  $\mu$  with probability at least  $(1 - 1/n)$ , if it finds an index  $i$  in  $(S_c \cap \tau)$ . So to prove the correctness of the algorithm, it is enough to show that with probability at least  $1/3$ ,  $(S_c \cap \tau) \neq \emptyset$ .

Now  $S_c$  has  $4n/2^c$  number of elements chosen uniformly at random from  $[n]$ . So probability that none of the elements of  $\tau$  is in  $S_c$  is

$$\left(1 - \frac{|\tau|}{n}\right)^{\frac{4n}{2^c}}.$$

Since  $|\tau| \geq 2^{c-1}$  the probability that  $S_c \cap \tau = \emptyset$  is less than  $(1/e)^2 \leq 1/3$ .

This prove the soundness of the algorithm.

### A.4 Proof of Theorem 5.1

**Lemma A.4.** [9] Given  $0 < \epsilon < 1$  and  $0 < \delta < 1$  and distributions  $\mu, \mu'$  over  $\Sigma$  there is an  $(\epsilon, \delta)$ -Identity Tester with conditional sample complexity  $\tilde{O}(\frac{\log \log |\Sigma|}{\epsilon^5} \log(\frac{1}{\delta}))$ . In other words, there is a tester that draws  $\tilde{O}(\frac{\log \log |\Sigma|}{\epsilon^5} \log(\frac{1}{\delta}))$  number of independent conditional samples and



- if  $\mu = \mu'$ , then the tester will accept with probability  $(1 - \delta)$ , and
- if  $d(\mu, \mu') \geq \epsilon$  then the tester will reject with probability  $(1 - \delta)$ .

The following algorithm takes (conditional) oracle access to  $\mu$ , and a threshold  $\epsilon$ . The algorithm accepts if  $\mu$  is a product distribution and rejects if  $\mu$  is  $\epsilon$  far from the product of marginals. Algorithm 3 uses **BasicUnknown**, the identity tester between two unknown distribution guaranteed in Lemma A.4 as a subroutine.

---

**Algorithm 3** The Independence Tester of Marginals

---

```

1:  $\epsilon' = \epsilon/2^{j+1}H(n)$ 
2:  $\delta' = \frac{\delta\epsilon}{24n(\log n)^2}$ 
3: for  $j = 1$  to  $\log n + 1$  do
4:   Create a set  $S_j$  by sampling, with replacement,  $(4n/2^j)$  element from  $[n]$  uniformly at random.
5:   for all  $i \in S_j$  do
6:     for  $k' = 1$  to  $3/\epsilon'$  do
7:       Sample  $w \sim \mu$ . Let  $w = (w_1, \dots, w_n)$ .
8:       if BasicUnknown $(\mu_i|w^{(i-1)}, \mu_i, \epsilon', \delta')$  rejects then
9:         Output REJECT
10:      end if
11:    end for
12:  end for
13: end for
14: Output ACCEPT

```

---

**Simulating Oracle Queries**

We first show that, given an access to a subcube conditional sampling oracle of  $\mu$ , we can correctly respond to the oracle queries made by **BasicUnknown**. To answer the conditional queries with condition  $B \subseteq \Sigma$  for the distribution  $\mu_i|w^{(i-1)}$ , we set  $A_j = \{w_j\}$  for  $j = 1, 2, \dots, i-1$ ,  $A_i = B$ , and  $A_j = \Sigma$  for  $j = i+1, \dots, n$ . We query the subcube conditional oracle for  $\mu$  with the condition  $A = A_1 \times \dots \times A_n$ . We respond with the sample that the subcube conditional oracle returns. Similarly, to answer conditional query for  $\mu_i$ , we set  $A_i = B$  and  $A_j = \Sigma$  for all  $j \neq i \in [n]$ .

**Sample Complexity of Algorithm 3**

By Lemma A.3, each invocation of **BasicUnknown** with parameter  $\epsilon', \delta'$  requires  $\tilde{O}(\log \log |\Sigma|/\epsilon'^5) = \tilde{O}\left(\frac{2^{5(j+1)}(H(n))^5 \log \log |\Sigma|}{\epsilon^5}\right)$  samples. For each index in  $S_j$ , **BasicUnknown** is queried  $3 \cdot 2^{j+1}H(n)/\epsilon$  times. Hence, the total sample complexity of our independence tester is

$$\begin{aligned}
\sum_{j=1}^{\log n+1} \frac{4n}{2^j} \times \frac{3 \cdot 2^{j+1}H(n)}{\epsilon} \times \tilde{O}\left(\frac{2^{5j+5}H(n)^5 \log \log |\Sigma|}{\epsilon^5}\right) &= \tilde{O}\left(\frac{nH(n)^5 \log \log |\Sigma|}{\epsilon^5}\right) \sum_{j=1}^{\log n+1} 2^{5j} \\
&= \tilde{O}\left(\frac{n^6 \log \log |\Sigma|}{\epsilon^5}\right)
\end{aligned}$$

### Correctness of Algorithm 3

To prove completeness, we need to show that if  $\mu$  is a product distribution, then Algorithm 3 accepts with probability at least  $2/3$ . For soundness, we need to prove that, if  $\mu$  is  $\epsilon$  far from *all* product distributions, the algorithm rejects with probability at least  $2/3$ . The proof is exactly same as for Algorithm 1.

We stress that, we do not need to consider a joint distribution  $\mu$  which is  $\epsilon$ -far from the product of its marginals but possibly  $\epsilon$ -close to some product distribution. Such a  $\mu$  does not satisfy the required distance condition required for soundness.  $\square$

## B Leftout Proofs for the lower Bound

### B.1 Proof of Claim 6.3

Let  $\mu = \mu_1 \otimes \cdots \otimes \mu_n$ . Without loss of generality we will assume that all the  $\mu_i$ 's are the distribution  $D_1$ . That is 1 is produced with probability  $(1/2 + 2\sqrt{\frac{\epsilon}{n}})$  and 0 produced with probability  $(1/2 - 2\sqrt{\frac{\epsilon}{n}})$ . For simplifying notations we will assume 1 is produced with probability  $(1/2 + \epsilon')$  and 0 produced with probability  $(1/2 - \epsilon')$ .

Since we know  $d(\mu, \mathcal{U}) \geq H(\mu, \mathcal{U})^2$ , it is enough for us to prove  $H(\mu, \mathcal{U})^2 \geq \epsilon$ . For any  $x \in \{0, 1\}^n$  let  $p(x)$  be the probability of getting  $x$  when drawn from  $\mu$ . Note that the probability of getting  $x$  when drawn from  $\mathcal{U}$  is  $1/2^n$ .

By definition we have

$$H(\mu, \mathcal{U})^2 = \frac{1}{2} \sum_{x \in \{0, 1\}^n} \left( \sqrt{p(x)} - \sqrt{1/2^n} \right)^2 = 1 - \sum_{x \in \{0, 1\}^n} \left( \sqrt{p(x)/2^n} \right)$$

Now note that if  $x$  has  $k$  1's and  $(n - k)$  0's then  $p(x) = (1/2 + \epsilon')^k (1/2 - \epsilon')^{n-k}$ . So we have

$$\sum_{x \in \{0, 1\}^n} \left( \sqrt{p(x)/2^n} \right) = \frac{1}{2^n} \sum_{k=0}^n \binom{n}{k} \sqrt{(1 + 2\epsilon')^k (1 - 2\epsilon')^{n-k}} = \frac{1}{2^n} \left( \sqrt{(1 + 2\epsilon')} + \sqrt{(1 - 2\epsilon')} \right)^n$$

Now since  $(\sqrt{1+x} + \sqrt{1-x}) \leq 2(1 - \frac{x^2}{8})$  for all  $x \leq 1$  so,

$$\frac{1}{2^n} \left( \sqrt{(1 + 2\epsilon')} + \sqrt{(1 - 2\epsilon')} \right)^n \leq \left( 1 - \frac{\epsilon'^2}{2} \right)^n \leq \left( 1 - \frac{\epsilon'^2 n}{2} + \frac{\epsilon'^4}{4} \binom{n}{2} \right).$$

The last inequality follows from the fact that  $(1 - x)^n \leq (1 - xn + \binom{n}{2}x^2)$ . Now putting all the things together we have

$$H(\mu, \mathcal{U})^2 \geq \left( 1 - \left( 1 - \frac{\epsilon'^2 n}{2} + \frac{\epsilon'^4}{4} \binom{n}{2} \right) \right) \geq \left( \frac{\epsilon'^2 n}{2} - \frac{\epsilon'^4}{4} \binom{n}{2} \right)$$

If  $\epsilon' = 2\sqrt{\epsilon/n}$  then from the above inequality we have  $H(\mu, \mathcal{U})^2 \geq \epsilon$ .

### B.2 Proof of Lemma 6.4

Let us start with a Claim. We defer the proof of the Claim to the end of this Section.

**Claim B.1.** *If  $P$  and  $Q$  be two distributions over  $\Sigma$  and for all  $x \in \Sigma$  we have*

$$\Pr_P(x) = (1 + \epsilon_x) \Pr_Q(x)$$

then we have

$$H(P, Q)^2 \leq \frac{1}{2} \sum_{x \in \Sigma} \epsilon_x^2 \Pr_Q(x)$$

Claim B.1 helps to upper bound the Hellinger distance in terms of the  $\ell_\infty$  distance. Now let  $\Sigma = \{0, 1\}^q$ . And let  $\mu_i^q$  be the distribution on  $\Sigma$  that is obtained by drawing  $q$  samples from  $\mu_i$ . Clearly,  $\mu^q = \mu_1^q \otimes \mu_2^q \otimes \dots \otimes \mu_n^q$ . To prove that the variation distance of  $\mu^q$  from uniform is less than  $1/3$  we will first show that the  $\ell_\infty$  distance of  $\mu_i$  from uniform is small then using Claim B.1 we get that the Hellinger distance of  $\mu_i^q$  from uniform is small.. And then we can show that if all the  $\mu_i^q$  has small Hellinger distance from uniform then  $\mu^q$  has small Hellinger distance from uniform which would give an upper bound on the variation distance of  $\mu^q$  from uniform.

Now the following Claim upper bounds the  $\ell_\infty$  distance of  $\mu_i^q$  from uniform.

**Claim B.2.** For all  $i$  and for all  $x \in \Sigma$

$$|\Pr_{\mathcal{U}}(x) - \Pr_{\mu_i^q}(x)| \leq \frac{10\epsilon q^2}{2qn}$$

Or in other words for all  $x \in \Sigma$  if

$$\Pr_{\mu_i^q}(x) = (1 \pm \epsilon_x) \Pr_{\mathcal{U}}(x)$$

then  $|\epsilon_x| \leq 10\epsilon q^2/n$

By definition of Hellinger distance and variation distance we have

$$d(\mu^q, \mathcal{U}) = \sum_{x \in \{0,1\}^{qn}} \left| \Pr_{\mu^q}(x) - \Pr_{\mathcal{U}}(x) \right| \leq 2H(\mu^q, \mathcal{U})$$

Again we know that for any two product distributions  $P = P_1 \otimes \dots \otimes P_n$  and  $Q = Q_1 \otimes \dots \otimes Q_n$

$$H(P_1 \otimes \dots \otimes P_n, Q_1 \otimes \dots \otimes Q_n)^2 \leq \sum_{i=1}^n H(P_i, Q_i)^2.$$

Thus we have

$$d(\mu^q, \mathcal{U}^q) \leq 2 \sqrt{\left( \sum_{i=1}^n H(\mu_i^q, \mathcal{U})^2 \right)} \quad (1)$$

From Equation 1 and Claim B.1 we have

$$d(\mu^q, \mathcal{U}^q) \leq 2 \sqrt{\sum_{i=1}^n \frac{1}{2} \sum_{x \in \Sigma} q(x) \epsilon_x^2},$$

where,  $q(x) = \Pr_{\mathcal{U}}(x)$ . So  $q(x) = 2^{-q}$ . From Claim B.2 we have that  $\epsilon_x = 10\epsilon q^2/n$ . So we have

$$d(\mu^q, \mathcal{U}) \leq 2 \sqrt{\sum_{i=1}^n (10\epsilon q^2/n)^2}$$

Thus if  $q \leq \sqrt[4]{n}/20\sqrt{\epsilon}$  we have  $d(\mu^q, \mathcal{U}) \leq 2\sqrt{1/40}$  which is less than  $1/3$

### B.2.1 Proof of Claim B.1

Let  $p(x) = \Pr_P(x)$  and  $q(x) = \Pr_Q(x)$ . By definition

$$H(P, Q)^2 = \frac{1}{2} \sum_{x \in \Sigma} \left( \sqrt{p(x)} - \sqrt{q(x)} \right)^2 = \left( 1 - \sum_{x \in \Sigma} \sqrt{p(x)q(x)} \right)$$

Now  $\sqrt{p(x)q(x)} = q(x)\sqrt{1 + \epsilon_x}$ . Now it is easy to verify that for all  $x$  such that  $|x| \leq 1$  we have

$$\sqrt{1 + x} \geq 1 + \frac{x}{2} - \frac{x^2}{2}$$

So, from the above observation we have,

$$\sqrt{p(x)q(x)} = q(x)\sqrt{1 + \epsilon_x} \geq q(x) \left( 1 + \frac{\epsilon_x}{2} - \frac{\epsilon_x^2}{2} \right)$$

Now since  $\sum_x q(x) = 1$  and  $\sum_x q_x \epsilon_x = 0$  so we have

$$H(P, Q)^2 \leq \left( 1 - \sum_x q(x) \left( 1 + \frac{\epsilon_x}{2} - \frac{\epsilon_x^2}{2} \right) \right) = \frac{1}{2} \sum_{x \in \Sigma} q(x) \epsilon_x^2$$

### B.2.2 Proof of Claim B.2

Let  $x \in \Sigma$  has  $k$  1's and  $(q - k)$  0's. Since the  $\mu_i$  is either the distribution  $D_1$  with probability  $1/2$  or distribution  $D_2$  with probability  $1/2$ , so probability of  $x$  appearing when drawn from  $\mu_i^q$  is

$$\frac{1}{2} \left( \left( \frac{1}{2} + \epsilon' \right)^k \left( \frac{1}{2} - \epsilon' \right)^{q-k} + \left( \frac{1}{2} - \epsilon' \right)^k \left( \frac{1}{2} + \epsilon' \right)^{q-k} \right) = \frac{1}{2^q} \frac{(1 + 2\epsilon')^k (1 - 2\epsilon')^{q-k} + (1 - 2\epsilon')^k (1 + 2\epsilon')^{q-k}}{2}$$

Note that since  $(1 + x)^n \geq 1 + xn$  so we have

$$(1 + 2\epsilon')^k (1 - 2\epsilon')^{q-k} + (1 - 2\epsilon')^k (1 + 2\epsilon')^{q-k} \geq (1 + 2k\epsilon')(1 - 2(q - k)\epsilon') + (1 - 2k\epsilon')(1 + 2(q - k)\epsilon')$$

The right hand side is equal to  $(2 - 8k(q - k)\epsilon'^2)$ . Thus we have

$$\Pr_{\mu_i^q}(x) \geq \frac{1}{2^q} (1 - 4k(q - k)\epsilon'^2) \geq \frac{1}{2^q} \left( 1 - \frac{4q^2\epsilon}{n} \right)$$

Now for the upper bound we use the inequality  $(1 + x)^n \leq 1 + xn + n^2x$ . So

$$(1 + 2\epsilon')^k (1 - 2\epsilon')^{q-k} \leq (1 + 2k\epsilon' + 4k^2\epsilon'^2)(1 - 2(q - k)k\epsilon' + 4(q - k)^2\epsilon'^2),$$

$$(1 - 2\epsilon')^k (1 + 2\epsilon')^{q-k} \leq (1 - 2k\epsilon' + 4k^2\epsilon'^2)(1 + 2(q - k)k\epsilon' + 4(q - k)^2\epsilon'^2)$$

and thus

$$\frac{(1 + 2\epsilon')^k (1 - 2\epsilon')^{q-k} + (1 - 2\epsilon')^k (1 + 2\epsilon')^{q-k}}{2} \leq (1 + 2\epsilon'^2 q^2 + q^4 \epsilon'^4)$$

Since  $\epsilon' = 2\sqrt{\epsilon/n}$  and  $q \leq \sqrt[4]{n}$  so we have

$$(1 + 2\epsilon'^2 q^2 + q^4 \epsilon'^4) \leq \left( 1 + \frac{10\epsilon q^2}{n} \right)$$

And thus we have

$$\frac{1}{2^q} \left( 1 - \frac{4q^2\epsilon}{n} \right) \leq \Pr_{x \leftarrow \mu_i^q}(x) \leq \frac{1}{2^q} \left( 1 + \frac{10\epsilon q^2}{n} \right)$$